19CSE304 – Assignment 3

CB.EN.U4CSE20216

Problem:

In 2010, the ACLU presented a report on jury selection in Alameda County, California. Therefore concluded that certain ethnic groups are underrepresented among jury panelists in Alameda County, and suggested some reforms of the process by which eligible jurors are assigned to panels. In this section, we will perform our own analysis of the data and examine some questions that arise as a result.

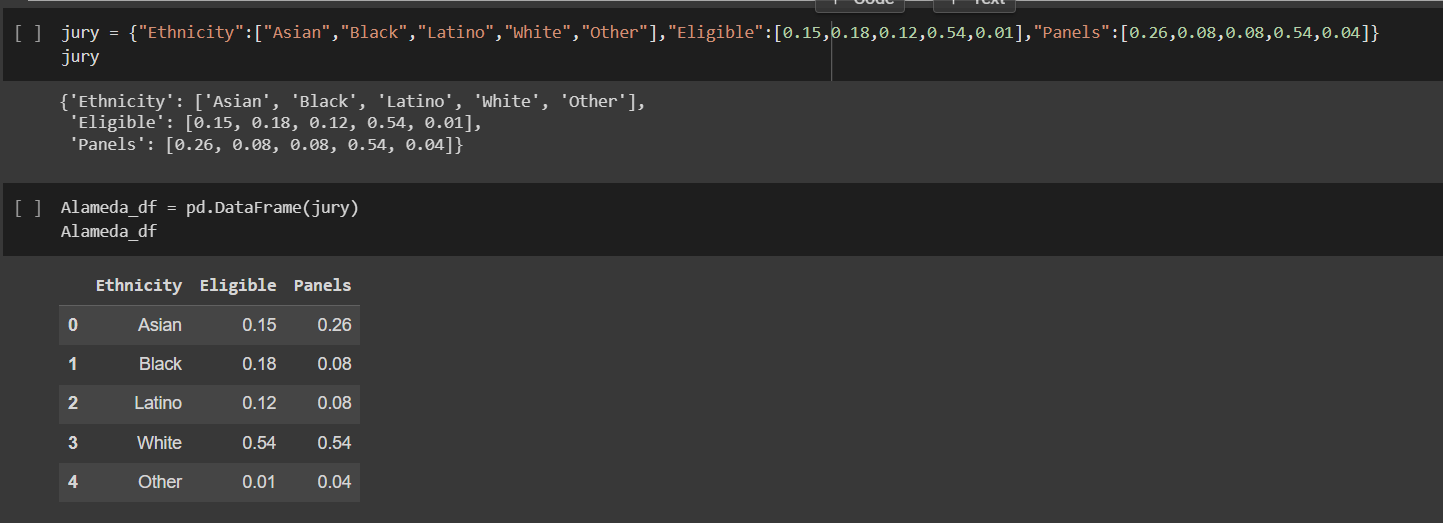
Attributes:

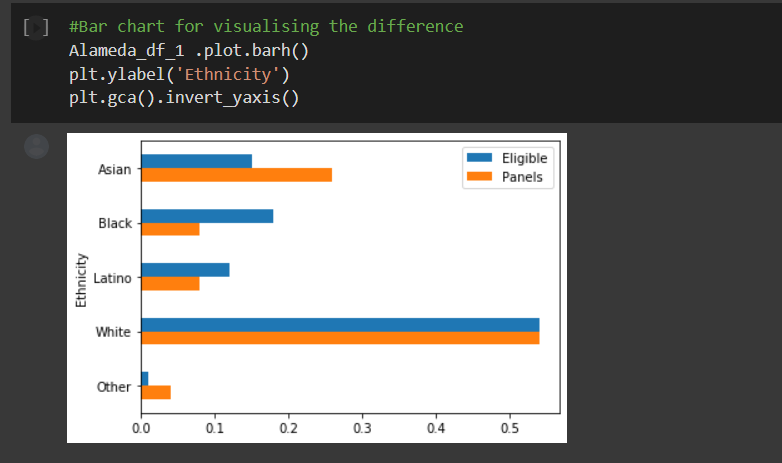
* **Ethnicity** tells us about the ethnic categories of people in Alameda.
* **Eligible** tells us about the percentage of people eligible grouped by ethnicity
* **Panel** tells us about the proportion of people currently chosen for the panel.

Hypothesis:

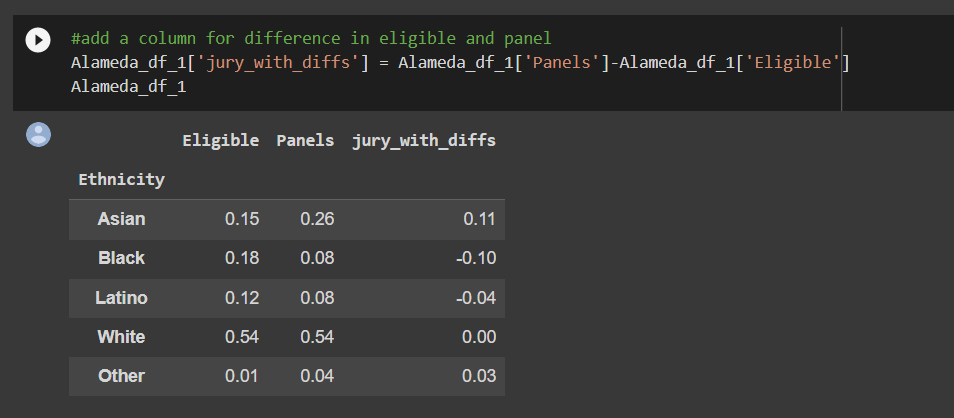
##### Null Hypothesis: - Panels were selected at random from the population of eligible jurors.

Alternate Hypothesis: - Panels were not selected at random.

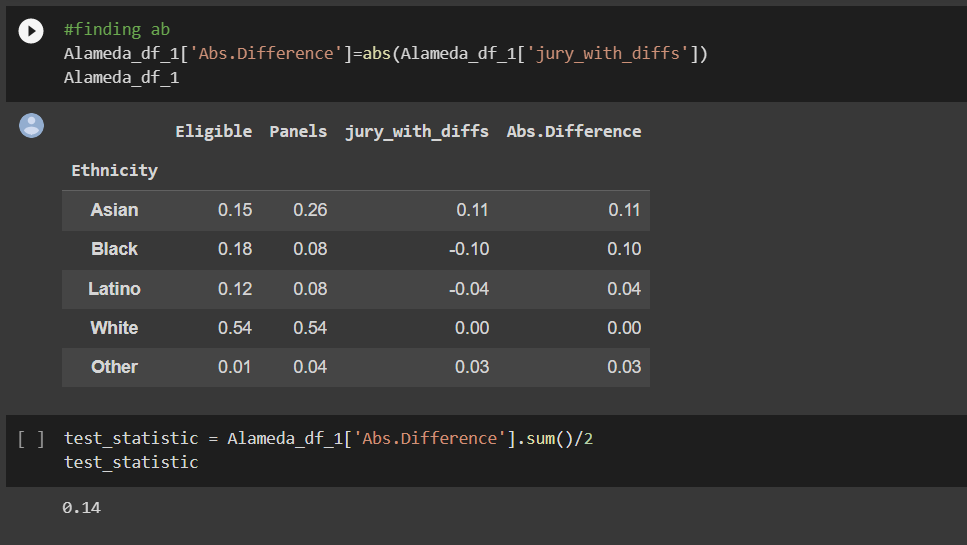




From the above bar graph, we can analyze that the eligible and panel members are not equal for the majority of the ethnicity. So, we can see the difference between them.

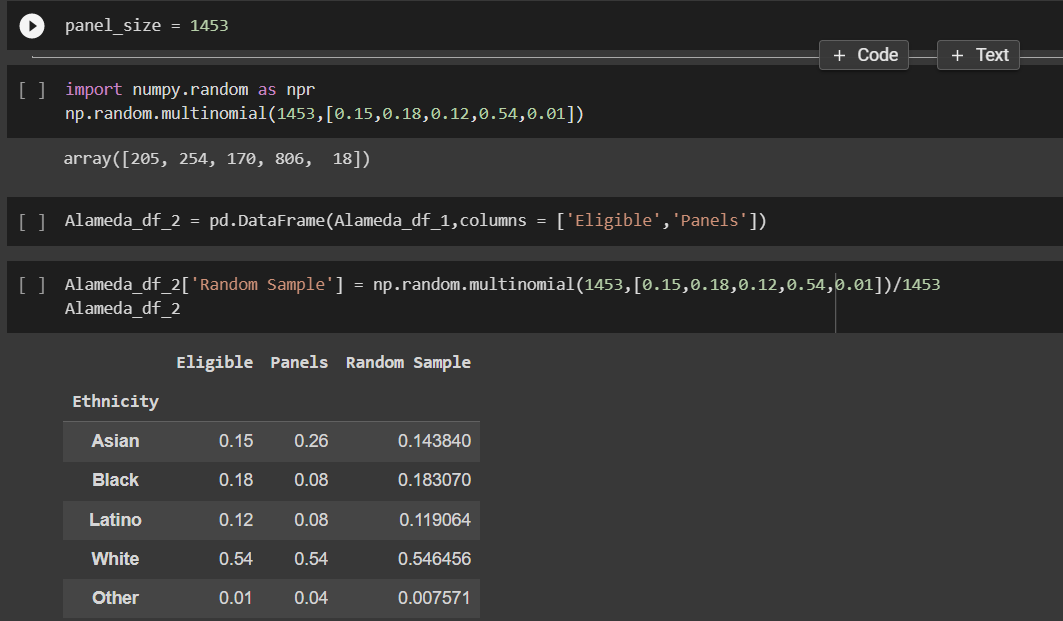


Take a look at the column Difference and notice that the sum of its entries is 0: the positive entries add up to 0.14, exactly canceling the total of the negative entries which is -0.14.

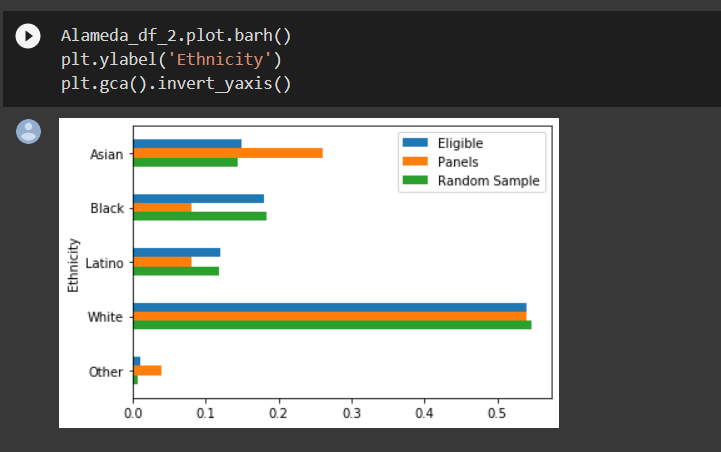


This is numerical evidence of the fact that in the bar chart, the gold bars exceed the blue bars by exactly as much as the blue bars exceed the gold. The proportions in each of the two columns Panels and Eligible add up to 1, and so the give-and-take between their entries must add up to 0.

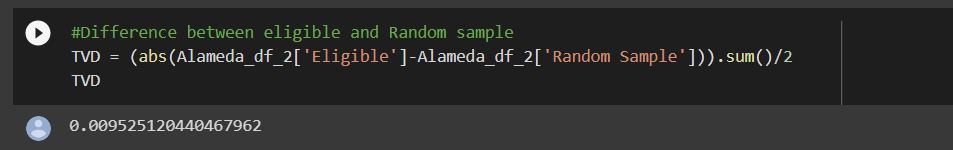
To avoid the cancellation, we drop the negative signs and then add all the entries. But this gives us two times the total of the positive entries (equivalently, two times the total of the negative entries, with the sign removed). So we divide the sum by 2.

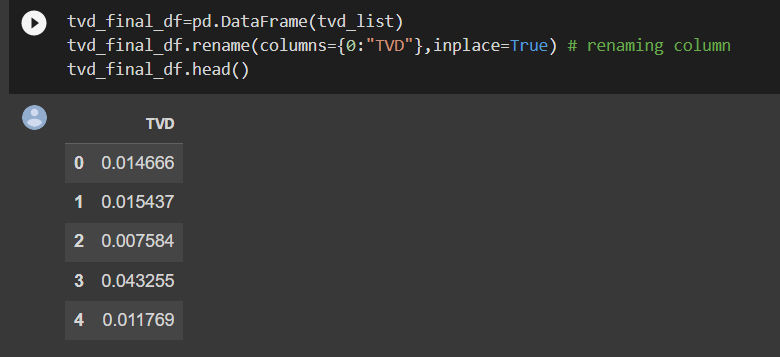


The distribution of the random sample is close to the distribution of the eligible population and is different from the distribution of the panels.

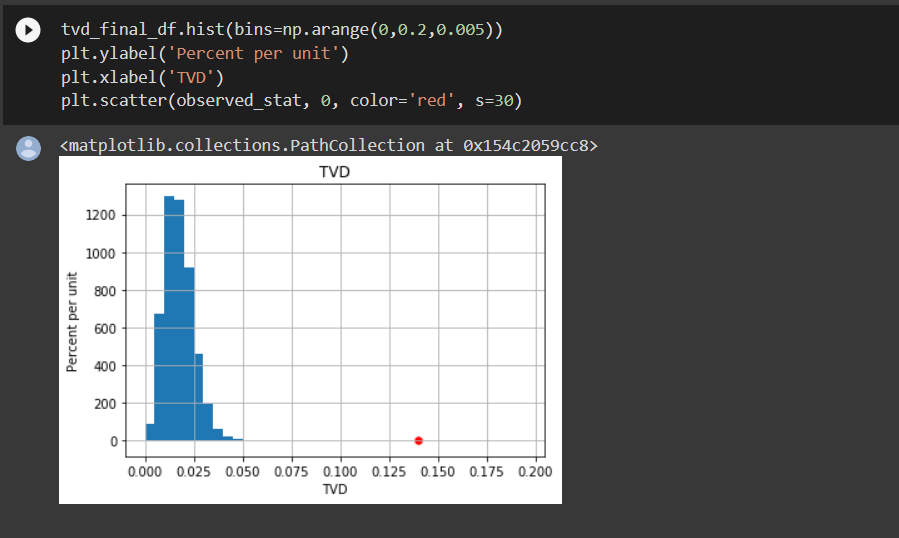


The Total variation distance(TVD) between the eligible population and the panels was 0.14, but the TVD between the eligible population and the random sample is much smaller.





Each row of the column above contains the total variation distance between a random sample of size 1453 (the size of the panels) and the population of eligible jurors.



The distance between the panels and the population was, so it is far away from the distribution. panels were not representative of the distribution provided for the eligible jurors.

Hence Null hypothesis is rejected